DOCUMENT RESUME

ED 454 895 JC 010 478

AUTHOR Armstrong, William B.
TITLE Pre-Enrollment Placement Testing and Curricular Content: Correspondence or Misalignment?
PUB DATE 2001-01-00
NOTE 12p.
PUB TYPE Reports - Research (143)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Tests; *Community Colleges; *Curriculum Based Assessment; *Educational Testing; *Student Placement; Two Year Colleges

ABSTRACT

This is a study of the relationship between placement test scores and academic achievement, as measured by the gain in placement pre- and post-test scores after students completed a semester of English instruction. Two placement tests were administered to a cohort of students enrolling in community college English courses. Pre- and post-placement test scores were analyzed for this cohort of students to determine the impact of instruction on post-test scores. No significant differences were found in pre- and post-test placement scores after a semester of English instruction for both the writing and reading placement tests. This investigation highlighted the problem of curricular misalignment with the placement tests used to sort and classify students on the basis of ability and aptitude for success in college English courses. The discussion section focuses on the fundamental problem of pre-enrollment placement testing that displays little or no relationship to course content, but a strong relationship to a post-test, administered after the term. A primary recommendation of this investigation is that colleges using placement tests to sort and classify students for the purposes of placement pay strict attention to aligning placement test content with instructional objectives of the course. (JA)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 454 895

Pre-Enrollment Placement Testing and Curricular Content: Correspondence or Misalignment?

William B. Armstrong, Ph.D.

University of California, San Diego

Abstract

This is a study of the relationship between placement test scores and academic achievement as measured by the gain in placement pre- and post-test scores after students completed a semester of English instruction. Two placement tests were administered to a cohort of students enrolling in community college English courses. These tests were the Conventions of Written English and the Reading Comprehension Test. Both placement tests are sub-tests of Descriptive Test of Language Skills (DTLS) battery developed by the College Board for use with community college students. The placement tests were administered at the start of the fall and subsequent spring semester. Analysis of the data revealed a cohort of 90 students who had taken the DTLS Reading Comprehension test in the fall, and also sat for the same reading test at the start of the next spring semester. Analysis of the cohort of students also showed that almost 80 students had taken the DTLS Conventions of Written English test at the start of the fall term also took the same test again at the start of the spring term. These findings enabled the study of the change in test scores after students completed a semester of instruction in English. Pre- and post- placement test scores were analyzed for this cohort of students to determine the impact of instruction on post-test scores. Matched pairs t-test analysis found no significant differences in pre- and post-test placement scores after a semester of English instruction for both the writing and reading placement tests. This finding of no significant gains in post-test scores was also found for students who completed a lower level English course, passed successfully, and enrolled in the next level English course in the next term. Additional analysis found that the correlation coefficient between the pre- and post-test scores was statistically significant, while the correlation between placement test score and course grade was not statistically significant. This investigation highlighted the problem of curricular misalignment with the placement tests used to sort and classify students on the basis of ability and aptitude for success in college English courses. The discussion section focused on the fundamental problem of pre-enrollment placement testing that displays little or no relationship to course content, but a strong relationship to a post-test, administered after the term. A primary recommendation of this investigation is that colleges using placement tests to sort and classify students for the purposes of placement, pay strict attention to aligning placement test content with instructional objectives of the course. This is particularly important in community colleges where mandatory placement rules are enforced. Aligning the instructional objectives of the course, with the skills assessed on placement tests, will help to ensure the integrity of the curriculum, and fairness for students.

RC010478

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

W. Armstrong

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

Introduction

In California and across the nation, there has been increasing interest and debate over the use of standardized tests for student access, placement, and certification. This has been particularly true in the community colleges where tests are used for placing students into curricular levels, determining eligibility for high-demand vocational programs, to determine mastery of course and program content, and to document institutional accountability.

At the same time, there is also a growing tension between the proponents of testing and those who view testing with suspicion (Armstrong, 2000)[1]. In 1997, education scholar, Michael Kirst[2] observed that in California higher education the lack of standardization in admissions and placement testing combined with a failure to behaviorally link tests with state curriculum guidelines creates problems for both students and schools. Kirst argues strongly for better alignment of placement tests and criteria used in the three segments of higher education in California with the California K-12 assessment tests and state curriculum guidelines. Test and curricular alignment may help to improve the correspondence between what is taught in classes, and what is measured on tests. Test and curricular misalignment is also a problem faced in the community colleges when standardized placement tests are used to sort students into the pre-collegiate and college level curricula. Strict attention to aligning the content of a course with the skills measured by testing may help somewhat, but the lack of correspondence between what tests measure and what faculty teach may still contribute to low predictive validity coefficients.

---

[1] Armstrong, W.B (1999). Explaining Community College Outcomes by Analyzing Student Data and Instructor Effects. University of California, Los Angeles: Unpublished doctoral dissertation

[2] Kirst, M.W. (1997). Exam confusion: Admissions and placement tests lack standardization. Crosstalk. San Jose, CA: The California Higher Education Policy Center.

Placement Testing and Instructional Objectives

Many education scholars have long advocated the importance of aligning instructional objectives with the content found on placement, progress, and exit tests.[3] To the extent that the skills measured on placement tests are derived from the instructional objectives of the course of interest, then predictive validity should improve. On the other hand, if the placement test used to sort students according to curricular level does not reflect course content, then predictive validity is diminished. This issue of test and course correspondence is also related to questions about educational access and efficiency. If the tests used by institutions to sift and sort students according to aptitude have little in common with course objectives, then the integrity of the instructional program is threatened. There are also questions of fairness and educational equity. If institutions allocate opportunity on the basis of test scores, and those tests do not adequately reflect the skills needed for course success, then what are we measuring? To ensure the integrity and fairness of a placement system, many scholars have long argued for the congruence of course objectives and the content of placement tests used to sort students according to aptitude for success.

Learning Outcomes

This study also has implications for the much ballyhooed paradigm shift that is supposed to be taking place in the nation's two-year colleges. Proponents of the "learning paradigm" have stressed the need to refocus the community college mission from teaching institutions to learning institutions. [4] Proponents of the learning paradigm suggest that institutional accountability will be enhanced by focusing on learning outcomes rather than merely the provision of teaching and

---

[3] Popham, W.J. (1988). Educational evaluation, 2nd ed. New Jersey, Prentice Hall

[4] Barr, R. B. & Tagg, J. (1995). From teaching to learning--a new paradigm for undergraduate education. Change Magazine, 27 (6): 12-25.

training. . However, a problem arises with how institutions are to measure themselves as learning institutions to document and report learning gains for accountability and institutional improvement. According to O'Banion [5], to wholly implement a learning-driven system, the entire college will require reform, including: the measurement of units of learning based on knowledge instead of time spent in class. This measure of learning would likely include the development of tests designed to measure course mastery. Under the learning paradigm, colleges would be accountable for demonstrating student learning. This view of instructional accountability militates strongly for a close alignment of placement tests with course instructional objectives. The present study will help to place these issues in perspective and helps to illustrate some potential problems with the use of tests developed for one purpose (such as placement), and used as measures of learning.

Research Questions

This is a study of the relationship between placement test scores, student assessment, and academic achievement as measured by gain in placement post-test scores after students completed a semester of English instruction. This study was guided by a primary research question that asked: What is the effect on pre- and post-test placement test scores after a semester of English instruction? A secondary question research question focused on a comparison of the predictive validity coefficient between scores on two separate administrations of a placement test and the predictive validity coefficient between placement test and the criterion variable of final grade.

---

[5] O'Banion, T. (1996, August). Learning communities, learning organizations, and learning colleges. Leadership Abstracts. Mission Viejo, CA: League for Innovation in the Community College, 9(8).

Theoretical Model

Central to this study was the application of the tenets of Point-to-Point Theory (Asher and Sciarrino, 1974)[6]. Point-to-Point Theory suggests that the use of scores on a placement test to predict the often normative variable of final grade in a course will generally result in unreliable, low predictive validity coefficients. According to Asher and Sciarrino: "Information with the highest validity seems to have a point-to-point correspondence with the criterion." (p. 519). In their analysis Asher and Sciarrino demonstrated that work sample tests there were in effect miniature replicas of the actual job or task for which prediction of aptitude was desired consistently demonstrated higher validity coefficients. Knowledge of how a student has performed in similar settings and how the criterion for success was determined greatly improves the validity of selection and placement decisions. This should be particularly true if there is correspondence between the predictor and criterion space. Therefore the more shared elements between predictor and criterion, the stronger the correlation coefficient.

Study Participants

In anticipation of the conversion from the Assessment and Placement Services for Community Colleges (APS) to the Descriptive Test of Language Skills (DTLS), a field test of the DTLS over two semesters was conducted. At the conclusion of the field test, analysis of the data revealed that almost 90 students who had taken the DTLS Reading Comprehension test in the fall, also sat for the exam at the start of the spring semester. Also, the database showed that almost 80 students had taken the DTLS Conventions of Written English test in the fall and again at the start of the spring term. These findings enabled the study of the change in test scores after students completed a semester of instruction in English and possibly other subjects. The discovery of a pair

of pre- and post-test scores for approximately 85 subjects enabled analysis of test score changes for students at varying levels of English (basic skills, pre-collegiate, college level) and for those students who progressed from a lower level to a higher level of English.

The analysis used two groups of study participants. One group of students completed their English course and did not re-enroll in a higher level course. Further analysis was conducted with students who had taken a placement test at the start of the fall semester, completed an English course successfully (i.e., a grade of 'C' or higher) and enrolled in a higher level of English. This analysis was conducted to reduce the number of students who may have taken an English course but performed poorly, or were repeating the course. Approximately 60 students met these criteria and had pre- and post-test data for DTLS Reading test, while 46 students meeting these criteria had pre- and post-test data for DTLS Conventions of Written English test.

Method

The research questions guiding this study were made operational by two separate administrations of DTLS to a group of participants. Data from two versions of the DTLS were analyzed. These two tests were the DTLS Conventions of Written English and the DTLS Reading Comprehension test. The Conventions of Written English consists of 40 questions administered in 40 minutes. According to the test publisher this test is designed to determine how much the subject knows about using standard forms of written English. The DTLS Reading Comprehension test consists of 45 questions designed to be administered in 45 minutes. The reading comprehension test contains individual and sets of questions based on reading passages.

To provide evidence to answer research question one, pre- and post- placement test scores were analyzed for a cohort of students who took the DTLS placement tests in the fall term and again

---

[6] Asher, J.J. and Sciarrino, J. A. (1974). Realistic work sample tests: A review. Personnel Psychology. 27, 519-533

in the spring term. After a semester of instruction, scores on the pre- and post-tests were compared using a matched pairs t-test. Significance was set at the .05 level. Evidence for the secondary research question was gathered by examining the correlation coefficient between the two pairs of scores and comparing the coefficient with the correlation coefficient between test score and final grade in the English course.

Findings

The mean DTLS Reading test score for the fall term was 23.8 with standard deviation of approximately 7.8 and a standard error of .83. Eighty-nine students were retested at the start of the subsequent spring semester. The average score of this same group of students on the DTLS Reading test was 23.1 with a standard deviation of 6.8 and a standard error of .73.

The mean DTLS Conventions of Written English score for the fall cohort was 23.5 with a standard deviation of 6.7 and a standard error of. 75. Seventy eight participants in the original field test sample were retested in the spring term. The average score of the retested sample was 23.7 with a standard deviation of 6.4 and a standard error of .73. A paired samples t-test did not reveal a statistically significant difference between the pre- and post-test reading or writing scores. These data are shown in table 1 below.

Table 1: Average Test Scores of a Cohort of Students on Two Successive Administrations of the DTLS Reading and English Tests

Test	Mean Score Fall, 1997 (Std. Dev.)	Mean Score Spring, 1998 (Std. Dev.)	t value	Probability (2-tailed)	N (Pairs of Scores)
Reading	23.8 (7.8)	23.1 (6.8)	1.09	.28	89
Writing	23.5 (6.6)	23.7 (6.4)	-.42	.68	78

A similar comparison was conducted using participants who had pre- and post-test scores and who had successfully completed their English course in the fall term and enrolled in a higher-level English course in the spring semester. The results of this analysis are shown in table 2.

Table 2: Comparison of Placement Test Scores for Students Advancing One Level in the English Curriculum

Test	Mean Score Fall, 1997 (Std. Dev.)	Mean Score Spring, 1998 (Std. Dev.)	t value	Probability (2-tailed)	N (Pairs of Scores)
Reading	23.5 (7.3)	22.7 (6.2)	.93	.35	60
Writing	23.0 (5.3)	23.7 (5.4)	-.195	.06	46

This analysis also showed that even for students who passed their English course in the fall term with a grade 'C' or better and advanced to a higher level English course, t-test analysis did not reveal a statistically significant difference between pre- and post-test scores.

The secondary research question tested if scores on successive administrations of a placement test showed a greater point-to-point relationship (i.e., a higher correlation coefficient) than the coefficient between test score and final grade. The correlation of pre- and post-test scores for the 89 students taking the DTLS Reading test and for the 78 students taking the DTLS English test suggests that the pairs of placement test scores was statistically significant. Additionally, the correlation coefficient between the two scores was superior to the coefficient observed between test score and final course grade. This finding tends to support the merit of Point –to- Point theory in positing a relationship between similar measures. These results are shown in Table 3.

Table 3: Correlation of Pre- and Post -DTLS Reading and Writing Scores, and Two Measures of Final Course Grade in English Courses

Test	Reading Spring, 1998	English Spring, 1998	N (Pairs of Scores)	Final Grade (4 point scale)	Final Grade ('W' counted as 'F'	N
Reading Fall, 1997	.65*		89	.20*	.16*	2393
Writing Spring, 1997		.78*	78	.26*	.23*	2393

* $p<.05$

For this analysis, two measures of final grade were used. Both used grade notations converted into a traditional four-point scale. The first grade calculation excluded students who received a 'W' notation indicating that they withdrew from the course after the third week (A=4, B=3, C=2, D=1, and F=0). The second grade measure included the 'W' notation as a failing grade (A=4, B=3, C=2, D=1, and F and W=0). These two grade measures were used because prior analysis have suggested that students may withdraw late in the course because they are performing poorly, thus the withdrawal might be a proxy for a failing grade [7]. Also, the 'W' notation is approximately 15%-20% of all grade notations given during a typical semester at the community college where this study was conducted.

As expected, the correlation coefficient between the two test scores was found to be superior to the coefficient observed between test score and final course grade. The correlation coefficient between pre- and post-test scores for both the DTLS Reading and English tests exceeded the value observed for these same placement tests and final course grade. These results supported the tenets of Point-to-Point theory.

[7] Explaining community college outcomes by analyzing student data and instructor effects. Unpublished doctoral dissertation, University of California, Los Angeles.

Summary and Implications

The study focused on the problems of curricular misalignment with placement tests used in the community colleges. This analysis provided insights into the fundamental purposes of placement testing that displays little or no relationship to course performance yet shows a strong relationship to a post-test, administered after the term. This study was intended to contribute to the general debate over the merits and purposes of standardized placement testing in the community colleges. This study also suggests the importance of examining the goals of placement testing and the goals of an instructional sequence of courses with various levels of placement. More closely aligning a college's assessment and placement program with the desired behavioral objectives of a course would probably improve predictive validity and also improve the instructional outcomes for students (Cohen, 1970; 1987; Dubin and Taveggia, 1968)[8]. Matching assessment methods with curricular content would help students make more informed choices about their own academic abilities and aptitude. To the extent that college assessment practices are dissimilar to the actual skills needed for course success, then the assessment and placement goal of creating homogeneous student groupings for the purposes of instruction will continue to remain elusive.

The problem of student placement and predictive validity is complex. As noted by Cohen and Brawer (1987) in their analysis of the use of placement tests and predictive validity in the two-year colleges:

> Obtaining higher correlations is difficult, because the variables are inconstant. The psychometrists ask that the dependent variable, the course grades, be more reliable; the faculty seek tests that will predict student success regardless of the shifting criteria for grades. The result is that the faculty prefer to rely on their own measures, particularly of student writing skills. For obvious reasons, writing assessment is considered a better predictor of student grades in the English classes; the same staff who are marking the writing samples on the entrance examination are marking the

[8] Cohen, A.M. and Brawer, F.B. (1987). The collegiate function of the community college. San Francisco, CA: Jossey-Bass
Cohen, A.M. and Associates (1970). Objectives for college courses. Beverly Hills, CA: Glencoe Press.
Dubin, R. and Taveggia, T.C. (1968). The teaching-learning paradox: A comparative analysis of college teaching methods. Eugene, OR: Center for the Advanced Study of Educational Administration, University of Oregon.

writing assignments in class, while performance on a quick score test of word usage is at most analogous behavior.

Short of comprehensive exams that mirror the skills assessed on placement tests, there will always be a relatively low point-to-point correspondence between final grade and placement test scores. The question then focuses on: how much instructor grade point variation is tolerable? Another, perhaps more central question is if there appears to be little or no relationship between the skills are measured on tests, and the skills demanded in the classroom, the what knowledge is most worth teaching and possessing? If we are to use placement tests and cutting scores to determine student readiness for a course, yet after a semester of instruction, there is no change in the test scores, then is it equitable and fair for students to be placed and advised using standardized test scores. For the faculty, are tests that bear little or no relationship to the material emphasized in class useful for instructional purposes? There is no clear answer to these questions. These discussions will need to take place at colleges, districts, and perhaps with state policymakers.

The findings from this study point out the need for further research in this area. The results might be viewed as somewhat tentative given relatively small sample sizes and lack of random assignment to treatments. It is recommended that other colleges replicate this study with larger samples if possible to determine the reliability of the results found here.